Indonesian Journal of Electrical Engineering and Computer Science
Vol. 33, No. 3, March 2024, pp. 1966~1973
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v33.i3.pp1966-1973 ❒ 1966

Exploring public sentiments towards courier services in Malaysia through Twitter: a sentiment analysis approach

Mohamed Imran Mohamed Ariff[1], Rasyidin Muhammad Zaharin Raja Salim[1], Azilawati Azizan[1], Samsiah Ahmad[1], Noreen Izza Arshad[2], Mohammad Nasir Abdullah[3]

[1]Computing Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Perak Branch, Tapah Campus, Malaysia

[2]Department of Computer and Information Sciences, Faculty of Science and Information Technology, Universiti Teknologi Petronas, Seri Iskandar, Malaysia

[3]School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Perak Branch, Tapah Campus, Malaysia

Article Info

Article history:

Received Nov 27, 2023
Revised Jan 2, 2024
Accepted Jan 3, 2024

Keywords:

Courier services
Public perception
Public sentiments
Sentiment analysis
User opinions

ABSTRACT

This study delves into the sentiment analysis of courier services in Malaysia through the lens of tweets. With the escalating reliance on courier services for daily parcel deliveries, understanding public sentiment becomes imperative. Through a systematic methodology encompassing data gathering, preparation, alteration, analysis, modeling, and deployment, this study aims to furnish valuable insights and recommendations. A dynamic graphical representation of sentiment scores is presented for intuitive interpretation and informed decision-making. The study's significance lies in empowering Malaysians to make judicious choices regarding courier services, thereby fostering transparency in the industry. The study proposes several avenues for future enhancements, including the integration of autocorrect features, diversifying data sources and languages, incorporating deep learning techniques, and developing a user-friendly interface for enhanced visualization. This study serves as a guiding resource, illuminating sentiments surrounding courier services in Malaysia and paving the way for future advancements in sentiment analysis within this domain.

Corresponding Author:

Mohamed Imran Mohamed Ariff
Computing Sciences Studies, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA
Perak Branch, Tapah Campus, Malaysia
Email: moham588@uitm.edu.my

1. INTRODUCTION

Courier services has undeniably evolved into an inseparable component in the e-commerce landscape, as an increasing number of individuals, particularly the younger generation, embrace the unparalleled convenience and accessibility of the online shopping [1]. Further, according to Gulc [2], the pivotal role of e-commerce in the evolution of courier services is an indispensable factor shaping customer satisfaction and future objectives, as well as the triumph of e-retailers. The escalating prominence of online shopping, a primary catalyst for the growth of e-commerce, has emerged as a pivotal driver in the evolution of courier services in recent years. The courier service provision has transformed from its conventional door-to-door delivery model to a self-service paradigm that leverages non-human interfaces, facilitated by the integration of information technologies and contemporary logistics solutions [3]. Furthermore, the importance of courier services has been emphasized as they provide a bridge between intergration between e-commerce and customers.

Thus, it is imperative to regard courier services not merely as offerings by providers but as the outcomes of a collaborative partnership with customers. In companies that are oriented towards client satisfaction (i.e., courier services), customers should be seen as a valuable resource, providing essential insights into their experiences, requirements, and service quality expectations. Furthermore, as the competition in the courier services sector is undergoing a significant transformation in response to changing consumer preferences and increased competition, courier services companies are compelled to adapt to these shifts and prioritize customer satisfaction as a key driver for success [4], [5]. Companies are compelled to adapt to these shifts and prioritize customer satisfaction as a key driver for success [2], [6].

The expanding world wide web (WWW) has spurred customers to voice their thoughts, complaints, and satisfaction through social media platforms. Platforms like Twitter produce substantial volumes of opinion-based texts, known as tweets, which are widely accessible on the WWW [7]. Furthemore, it has been suggested by previous research the use of tweets can help companies better analyze their overall customer satisfaction as compared to the use of questionnaires as, collecting customer feedback through questionnaires can be challenging [8], [9]. Additionally, studies have suggested that Twitter generates a substantial volume of opinion-based text (i.e., tweets) which can be efficiently summarized and organized using sentiment analysis techniques [7], [10]. The utilization of sentiment analysis has emerged as an indispensable instrument for monitoring and comprehending the sentiments and emotions, thereby enhancing the understanding of customer satisfaction [11], [12].

2. LITERATURE REVIEW

Courier service is a type of logistics service that assists and provides parcel delivery for users (clients) who are selling or buying items online. The typical delivery process of any courier service is divided into 2 cycles either: i) morning or ii) afternoon. In Malaysia, courier service has contributed significantly to the national economy in terms of e-commerce sales volume [13]. However, despite the courier services contributing to the to the national economy, it has been widely reported in the social media platform the overall customer satisfaction with these courier service. Among the customer satisfaction which has been widely reported in the social media are: i) service quality, ii) delivery time, and iii) damage parcel delivery and experiences [7], [14], [15].

In recent years, there has been a substantial increase in the widespread adoption of social media for communication between customers and courier service companies, particularly in terms of how customers express their satisfaction concerns regarding the use of courier services and how these companies address such issues with equitable fairness [16]. Among the social media platforms, Twitter has emerged as the most extensively adopted platform in numerous business sectors, significantly impacting customer satisfaction [17], [18]. Twitter provides users with two primary benefits: i) a platform for seeking information and ii) a means for sharing information [5], [19]. These dual advantages enable both customers and courier service companies to promptly address concerns related to customer satisfaction, ultimately enhancing customer choices and the overall management of courier services by the companies [20]. Previous research has underscored sentiment analysis as the preeminent analytical technique for assessing opinions and reviews derived from social media [1], [21]. Sentiment analysis involves the meticulous examination of textual content to discern sentiment pertaining to a specific subject of interest [22]. Sentiment analysis is a pivotal process, delves into the textual content to discern the opinions related to a specific subject of interest [23]. It meticulously computes each text to ascertain its opinion tendency, categorizing it as either a positive or negative opinion tendency [24]. In this study, sentiment analysis is applied to discern customers positive and negative sentiments, a departure from prior research that employed sentiment analysis to gauge the quality of courier services through the use of data from Twitter [25], [26].

3. METHOD

In this study, the process of sentiment analysis pertains to Kaur *et al.* [27], as illustrated in Figure 1, which presents the research methodology. This methodology has six main stages, which are: i) gathering data, ii) data preparation, ii) data altercation, iv) analyze data, v) modelling, and vi) deployment is a total of six stages of the operational structure of this project. A machine learning life cycle (MLLC) is the method that used to evolve this project which will be a guideline to assist by ensuring that the development process is systematic. Each of the stages will be further explained in the sub-section.

3.1. Gathering data

In this study, Twitter was utilized as a data source due to its prominence in microblogging social media services, enabling users to convey concise messages within a 140-character limit per tweet [28]. Given Twitter's extensive user base, it serves as a significant resource for the research conducted in this study. During the data collection phase, valuable insights into public sentiments regarding courier services in Malaysia were easily obtained from the platform. This data assumes a pivotal role in subsequent modeling processes. The focus of this study extends to various courier service brands, including City-Link Express, DHL, FedEx, PosLaju, and Shopee Express. Each brand contributes a distinct set of data points to the overall dataset. The aggregation and integration of these diverse datasets result in the formation of a comprehensive dataset.

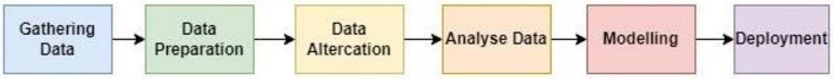

Figure 1. MLLC

3.2. Data preparation

In the second phase of MLCC, the primary objective is to prepare the data through randomization of its order and subsequent storage in a suitable repository. This initial data preparation step is integral to the subsequent phase of data exploration. Within this sub-phase, the ML model comprehensively grasps the inherent nature of the data, encompassing its characteristics, format, and quality. A profound understanding of these data attributes paves the way for effective outcomes. During this phase, correlations, general trends, and outliers are identified and scrutinized before proceeding to the data pre-processing stage. In the second sub-phase, data is pre-processed to facilitate subsequent analytical endeavor.

3.3. Data altercation

Data wrangling, commonly referred to as data alteration, involves the transformation of raw data into a more organized and presentable format suitable for analysis. This process is particularly crucial following the earlier data pre-processing phase. Given the likelihood of data being in tweet form, extracted from Twitter, data alteration becomes imperative due to the inherent difficulty in accessing and analyzing such data. Effective data alteration is vital for extracting critical insights that might otherwise remain obscured. The focus is on unstructured data, commonly found in digital information lacking a recognized data structure. Techniques employed include the cleaning and organization of unstructured data, entailing the removal of unwanted null data and the application of string processing methods to convert the data into a structured format. This transformation facilitates the identification of meaningful patterns and the generation of new insights [28]. The process also addresses missing data, as well as date and time considerations. Within this phase, the model undergoes a meticulous process of error identification and removal, targeting unwanted null values, characters, emojis, hashtags, URL links, and punctuations, as these elements adversely impact the predictive model [29], [30].

3.4. Analyze data

This phase involves developing an artificial intelligence (AI) ML model to analyze Twitter-derived tweet data, employing techniques such as stemming, tokenization, chunking, and lemmatization. The study focuses on sentiment analysis, extracting textual data from tweets to classify polarity. The chosen model, Naïve Bayes, excels in this task, surpassing similar models. The next step involves creating a ML model using Jupyter Notebook to train the Twitter-collected dataset.

3.5. Modeling

This phase encompasses two critical stages: dataset training and testing. It represents a pivotal juncture, wherein the processed data undergoes training and testing using the Naïve Bayes model. Model training involves inputting data into the ML algorithm to determine suitable values for project variables [31]. Conversely, model testing is imperative for verifying the accuracy of the trained model and assessing its effectiveness. Amidst various ML models, the project opts for the supervised method due to its effectiveness in training algorithms for accurate data categorization and outcome prediction. The model is trained with an 80-20 data split, with 80% used for training and the remaining 20% for testing to avoid biased data issues [32]. The algorithm trains the model by inputting data and comparing the processed output with sample output. This study focuses on analyzing reviews and public sentiments regarding courier services in Malaysia through tweets. To ascertain the accuracy of the prediction model, an accuracy score is computed using a formula, where higher values indicate better performance.

3.6. Deployment

During the deployment phase, meticulous efforts are invested in crafting a PowerBI dashboard designed to serve as the ultimate user interface. Simultaneously, the system's functionality undergoes rigorous and periodic testing to ensure optimal performance aligned with its intended purpose. Any issues identified during this testing phase are promptly addressed and resolved, emphasizing a commitment to maintaining the system at its highest capacity.

This proactive approach not only safeguards against potential disruptions but also contributes to delivering a top-notch user experience. The continuous monitoring and resolution of potential issues reflect the dedication to achieving and sustaining excellence in system performance throughout its operational lifecycle.

4. RESULTS AND ANALYSIS

This section adopts a structured approach organized around four key points: i) interface, ii) process of gathering tweets, iii) findings of sentiment analysis, and iv) evaluation of the sentiment analysis classifier model. Each of these critical components will be elaborated upon in the subsequent sub-sections, providing a comprehensive and detailed exploration of the study's results. The initial focus will be on delving into the intricacies of the interface, elucidating its design and functionality. Subsequently, the section will scrutinize the meticulous process involved in gathering tweets, shedding light on the methodology employed. Following this, the findings of sentiment analysis will be presented, offering valuable insights into the observed sentiment trends. Finally, a thorough examination of the sentiment analysis classifier model will be conducted, assessing its performance and overall efficacy. This structured and in-depth approach ensures a nuanced understanding of each aspect, contributing to a holistic interpretation of the study's outcomes

4.1. Interface

The creation of a web dashboard Figure 2 utilizing Tableau software is integral to this study, enhancing the visualization of analyzed data and thereby aiding in the identification of trends and patterns, crucial aspects of sentiment analysis. In a Tableau web dashboard, the unique capability to present multiple visualizations within a single view enhances data exploration and analysis. This consolidated approach provides users with a comprehensive overview, facilitating a more holistic understanding of the data by offering diverse insights in one centralized interface. The interactive nature of Tableau dashboards further empowers users to dynamically engage with the visualizations, enabling them to uncover patterns, correlations, and trends seamlessly.

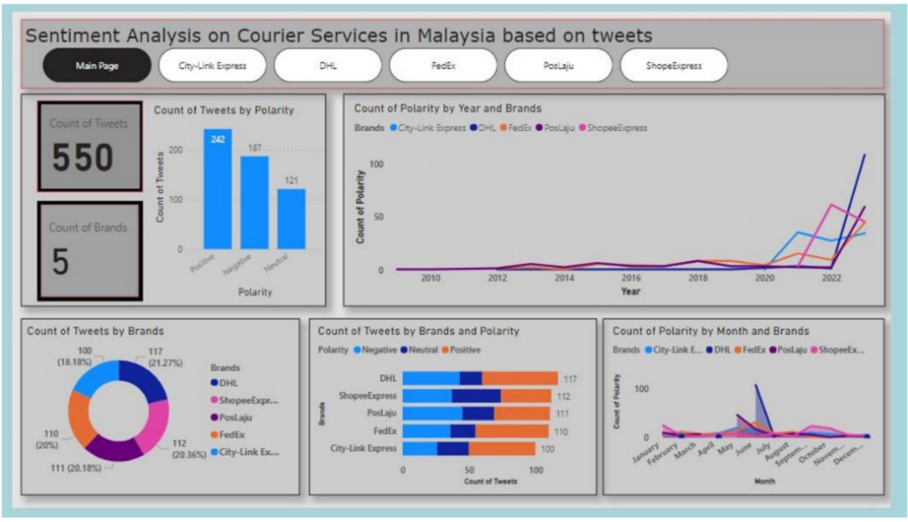

Figure 2. Web dashboard

4.2. Tweets gathering process

This section describes the process of collecting tweets for the dataset in this study. The Twitter data for this study was gathered manually using a normal user Twitter account. The manual collection of Twitter data for this study utilized a standard user account, employing an advanced search method. This method personalized the search query by incorporating specific phrases and words while excluding irrelevant terms for the dataset, through the application of specific language and time phase parameters. The required dataset was then manually generated using the search method, involving the extraction of details such as date, username, and tweets. Five prominent courier service brands in Malaysia-PosLaju, DHL, FedEx, Shopee Express, and City-Link Express-were selected for analysis, each contributing a minimum of one hundred datasets spanning various time phases. Subsequently, the dataset underwent data preparation and cleansing, as delineated in the methodology section. In the data cleaning phase, extraneous punctuation, stop words, and sentences were eliminated from the tweet column.

4.3. Sentiment analysis results

The examination of the refined dataset unveiled predominant neutrality in the tone of courier data, with only a limited portion exhibiting a negative sentiment, as depicted in Figure 3. This observation implies a prevailing positive disposition among individuals towards various courier companies in Malaysia, particularly in the context of customer satisfaction. This conclusion gains further support from an analysis of Google search reviews, which affirm the positive customer satisfaction associated with each specific courier company.

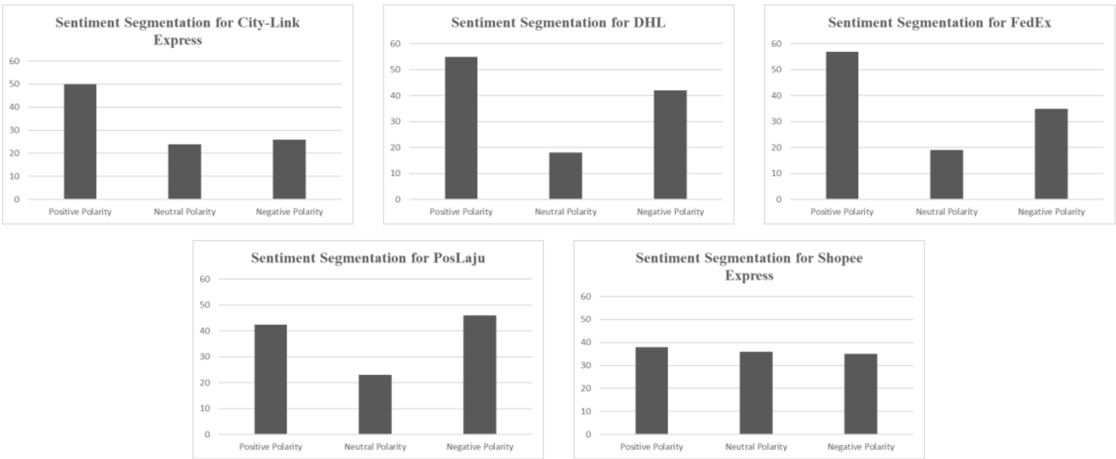

Figure 3. Views on courier satisfaction based on courier companies

In addition to the sentiment analysis, this study incorporates a word cloud to visually represent public opinions regarding courier companies. Figure 4 displays the word cloud, centering on topics associated with courier companies. Notably, the word cloud highlights Fedex and DHL as the most prominently discussed and tweeted entities on Twitter, emphasized by their larger font size and bolded letters.

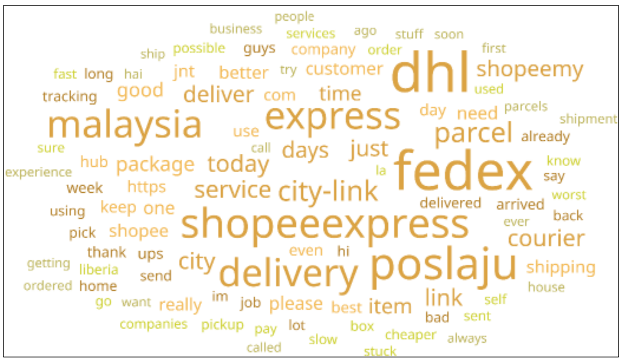

Figure 4. Word cloud outcome

4.4. Evaluation of the sentiment analysis model (classifier)

The evaluation of the sentiment analysis classifier involves the utilization of four standard metrics: i) precision, ii) recall, iii) F1-score, and iv) accuracy. These metrics are widely employed for assessing the performance of classification models. This study attained an accuracy rate of 90.3%, a commendable achievement given that any value surpassing 70% is considered indicative of a proficient model in the evaluation of sentiment analysis. Further, the precision score of this study is at 93%, which indicates the model effectively discerns genuine positive outcomes among all correctly predicted positive results. The recall score of this study was at 94% signifies the model's capacity to accurately predict every instance in the training dataset. Finally, the F1-score of 94% in this study further corroborates the model's efficacy and reliability in the realm of sentiment analysis.

5. CONCLUSION

The study aimed to provide a comprehensive understanding of Malaysians' perceptions about the courier services. To achieve this, data was collected from Twitter and analysed to determine the sentiment behind the tweets. The results of the study showes Malaysians have a generally neutral perception of the courier services in Malaysia. This study highlights the importance of understanding public perception and sentiment towards customer satisfaction in dealing with courier services. The findings can be used to inform courier companies in making informed decisions regarding customer satisfaction. This study can also be used as a foundation for future research in the field of sentiment analysis, with the potential for improvement and expansion. This study encountered limitations in its implementation, particularly in dealing with nuanced language features. The automated method struggles with interpreting sarcasm, irony, negations, jokes, and exaggerations, leading to potential inaccuracies in sentiment classification. For instance, the system may misclassify the word "sad" as negative, overlooking its positive context in phrases like "I was not sad". Similarly, detecting sarcasm poses a challenge, as exemplified by statements like "I'm really loving the enormous pool at my hotel!" accompanied by an image of a small pool. These challenges underscore the difficulty automated sentiment analysis tools face in accurately deciphering sentiment in complex language.

This study holds promise for further development and refinement. One avenue for improvement involves the implementation of real-time updates for the sentiment analysis results, enabling the continuous tracking of emotions among Malaysians. This eliminates the need for manual extraction of new tweets, enhancing user convenience. Additionally, incorporating a sentiment classifier correction tool is another potential enhancement. This tool could rectify typos and misspelled words in both the existing dataset and new tweets, thereby elevating data quality and increasing the overall accuracy of the classifier. Furthermore, there is potential to enhance user engagement by integrating multimedia elements such as photos and videos, along with additional navigation buttons and notifications. These features would keep users informed about the availability of data for analysis and evaluation. Looking ahead, the study aims to incorporate new algorithms to further enhance the accuracy of sentiment analysis, contributing to continuous advancements in this field of study.

ACKNOWLEDGEMENTS

This study was supported by the College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Perak Branch, Tapah Campus under the final year project (FYP) 2022 / 2023.

REFERENCES

[1] S. Loomba, M. Dave, H. Arolkar, and S. Sharma, "Sentiment analysis using dictionary-based lexicon approach: analysis on the opinion of Indian community for the topic of cryptocurrency," *Annals of Data Science*, 2023, doi: 10.1007/s40745-023-00496-y.
[2] A. Gulc, "Determinants of courier service quality in e-commerce from customers' perspective," *Quality Innovation Prosperity*, vol. 24, no. 2, Jul. 2020, doi: 10.12776/qip.v24i2.1438.
[3] M. Mazur, *Five forces transforming transport & logistics PwC CEE Transport & logistics trend book 2019*. PWC, 2019.
[4] J. T. Libo-on, "Service quality influence on customer satisfaction in courier services: a comparative study," *American International Journal of Business Management (AIJBM)*, vol. 4, no. 3, pp. 51–63, 2021.
[5] O. Chamorro-Atalaya *et al.*, "Sentiment analysis through twitter as a mechanism for assessing university satisfaction," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 1, pp. 430–440, Oct. 2022, doi: 10.11591/ijeecs.v28.i1.pp430-440.
[6] A. Gulc, "Courier service quality from the clients' perspective," *Engineering Management in Production and Services*, vol. 9, no. 1, pp. 36–45, Mar. 2017, doi: 10.1515/emj-2017-0004.
[7] M. A. Ligiarta and Y. Ruldeviyani, "Customer satisfaction analysis of mobile banking application based on Twitter data," in *2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, Nov. 2022, pp. 322–327. doi: 10.1109/ICE3IS56585.2022.10010221.
[8] Information Resources Management Association, *Social media marketing: Breakthroughs in research and practice*. IGI Global, 2018.
[9] R. Monika, S. Deivalakshmi, and B. Janet, "Sentiment analysis of US Airlines tweets using LSTM/RNN," in *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, Dec. 2019, pp. 92–95. doi: 10.1109/IACC48062.2019.8971592.
[10] A. Sarlan, C. Nadam, and S. Basri, "Twitter sentiment analysis," in *Proceedings of the 6th International Conference on Information Technology and Multimedia*, Nov. 2014, pp. 212–216. doi: 10.1109/ICIMU.2014.7066632.
[11] H. Najadat, A. Al-Abdi, and Y. Sayaheen, "Model-based sentiment analysis of customer satisfaction for the Jordanian telecommunication companies," in *2018 9th International Conference on Information and Communication Systems (ICICS)*, Apr. 2018, pp. 233–237. doi: 10.1109/IACS.2018.8355429.

[12] R. A. Laksono, K. R. Sungkono, R. Sarno, and C. S. Wahyuni, "Sentiment analysis of restaurant customer reviews on TripAdvisor using Naïve Bayes," in *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, Jul. 2019, pp. 49–54. doi: 10.1109/ICTS.2019.8850982.

[13] Suruhanjaya Komunikasi dan Multimedia Malaysia "2Q 2023 postal and courier services infographic," *Malaysian Communications and Multimedia Commission*, 2023. https://www.mcmc.gov.my/en/resources/statistics/postal-courier-pocket-book-of-statistics#(accessed November 25, 2023).

[14] N. A. Yaacob and T. Z. B. Yaacob, "Factors affecting the level of customer service satisfaction towards the courier service management quality at Johor Bahru," *Scientific Journal of Innovation and Social Sciences Research*, vol. 2, no. 1, pp. 27–44, 2022.

[15] N. F. B. Hasrunlah, N. F. B. Saidin, M. S. Bin Jalaluddin, Ka. Hu. Zhen, and J. Bin Jamsuri, "Customer satisfaction on parcel delivery company," Degree of Bachelor of Enterpreneurship (Logistics and Distributive Trade) with Honours, 2023.

[16] N. Arumugam, A. L. B. Subramaniam, G. Mylvaganam, and K. K. M. Singh, "Customer satisfaction perpetuates loyalty: a case study," *International Journal of Academic Research in Business and Social Sciences*, vol. 10, no. 12, Dec. 2020, doi: 10.6007/IJARBSS/v10-i12/8282.

[17] J. Bhattacharjya, A. Ellison, and S. Tripathi, "An exploration of logistics-related customer service provision on Twitter," *International Journal of Physical Distribution & Logistics Management*, vol. 46, no. 6/7, pp. 659–680, Jul. 2016, doi: 10.1108/IJPDLM-01-2015-0007.

[18] N. A. N. Mohd Idros, H. Mohamed, and R. Jenal, "The use of expert review in component development for customer satisfaction towards e-hailing," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 1, pp. 347–356, Jan. 2020, doi: 10.11591/ijeecs.v17.i1.pp347-356.

[19] S. Klar, Y. Krupnikov, J. B. Ryan, K. Searles, and Y. Shmargad, "Using social media to promote academic research: Identifying the benefits of twitter for sharing academic work," *PLOS ONE*, vol. 15, no. 4, Apr. 2020, doi: 10.1371/journal.pone.0229446.

[20] A. Wirapraja, H. Aribowo, and E. T. Setyoadi, "The influence of e-service quality, and customer satisfaction on go-send customer loyalty in Surabaya," *Indonesian Journal of Information Systems*, pp. 128–137, Feb. 2021, doi: 10.24002/ijis.v3i2.4191.

[21] J. N. Mindoro, M. A. F. Malbog, M. D. Nipas, J. A. B. Susa, A. G. Acoba, and J. S. Gulmatico, "Sentiment analysis in customer experience in Philippine courier delivery services using VADER algorithm thru chatbot interviews," in *2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T)*, Mar. 2022, pp. 1–6. doi: 10.1109/ICPC2T53885.2022.9777007.

[22] S. Bhuta and U. Doshi, "A review of techniques for sentiment analysis Of Twitter data," in *2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, Feb. 2014, pp. 583–591. doi: 10.1109/ICICICT.2014.6781346.

[23] E. T. Khalid, E. B. Talal, M. K. Khamees, and A. A. Yassin, "Sentiment analysis system for COVID-19 vaccinations using data of Twitter," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 2, pp. 1156–1164, May 2022, doi: 10.11591/ijeecs.v26.i2.pp1156-1164.

[24] E. Udayanti, A. Rahman, and E. Kartikadharma, "Sentiment analysis towards courier service: Case study on JNE Semarang," in *Proceedings of the The 1st International Conference on Computer Science and Engineering Technology Universitas Muria Kudus*, 2018. doi: 10.4108/eai.24-10-2018.2280499.

[25] R. Hendayani and M. C. Dharmawan, "Strategies for improving the quality of logistics courier services through priority problem-solving based on multiclass classification," *IOP Conference Series: Materials Science and Engineering*, vol. 879, no. 1, Jul. 2020, doi: 10.1088/1757-899X/879/1/012051.

[26] E. Sutoyo and A. Almaarif, "Twitter sentiment analysis of the relocation of Indonesia's capital city," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1620–1630, Aug. 2020, doi: 10.11591/eei.v9i4.2352.

[27] H. Kaur, V. Mangat, and Nidhi Nidhi, "A survey of sentiment analysis techniques," in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Feb. 2017, pp. 921–925. doi: 10.1109/I-SMAC.2017.8058315.

[28] A. Muneer and S. M. Fati, "A comparative analysis of machine learning techniques for cyberbullying detection on Twitter," *Future Internet*, vol. 12, no. 11, Oct. 2020, doi: 10.3390/fi12110187.

[29] J. Brownlee, "How to perform data cleaning for machine learning with python," *Machine Learning Mastery*, 2020. https://machinelearningmastery.com/basic-data-cleaning-for-machine-learning/ (accessed October 24, 2023).

[30] D. Uzun Ozsahin, M. Taiwo Mustapha, A. Saleh Mubarak, Z. Said Ameen, and B. Uzun, "Impact of outliers and dimensionality reduction on the performance of predictive models for medical disease diagnosis," in *2022 International Conference on Artificial Intelligence in Everything (AIE)*, Aug. 2022, pp. 79–86. doi: 10.1109/AIE57029.2022.00023.

[31] P. Assiroj, A. Kurnia, and S. Alam, "The performance of Naïve Bayes, support vector machine, and logistic regression on Indonesia immigration sentiment analysis," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 6, pp. 3843–3852, Dec. 2023, doi: 10.11591/eei.v12i6.5688.

[32] V. R. Joseph and A. Vakayil, "SPlit: an optimal method for data splitting," *Technometrics*, vol. 64, no. 2, pp. 166–176, Apr. 2022, doi: 10.1080/00401706.2021.1921037.

BIOGRAPHIES OF AUTHORS

Mohamed Imran Mohamed Ariff is an Associate Professor with the Computing Sciences Studies under the College of Computing, Informatics and Mathematics at Universiti Teknologi MARA (UiTM), Perak Branch, Tapah Campus, Malaysia. He has been with UiTM since 2003. His research interest includes, and not limited to: i) mobile applications, ii) internet of things (IoT), iii) sentiment analysis, iv) information systems, and v) smart-PLS. He can be contacted at email: moham588@uitm.edu.my.

Rasyidin Muhammad Zaharin Raja Salim is a dedicated final-year student currently pursuing a Bachelor's degree in Computer Science at the Universiti Teknologi MARA (UiTM), Perak Branch, Tapah Campus, Malaysia. Having recently concluded his industrial internship. Rasyidin now stands at a crossroads, contemplating whether to satiate his thirst for knowledge through the master's program or dive into the dynamic realm of technology. His research interests is firmly rooted in the captivating fields of sentiment analysis and machine learning. He can be contacted at email: rasyidin4714@gmail.com.

Dr. Azilawati Azizan is currently a senior lecturer at the College of Computing, Informatics and Mathematics at Universiti Teknologi MARA (UiTM), Perak Branch, Tapah Campus, Malaysia. She obtained her Ph.D. from UiTM Shah Alam Selangor in 2022. Her areas of interest are in text processing, query processing, and search engine. She can be contacted at email: azila899@uitm.edu.my.

Samsiah Ahmad obtained her first degree in Data Communication and Networking from the University Teknologi MARA, Malaysia in 2005. She is received a master’s degree from Universiti Teknologi Malaysia in Computer Science, with specialization in Information Security. Currently, she is a lecturer at the Computing Sciences Studies under the College of Computing, Informatics and Mathematics at Universiti Teknologi MARA (UiTM), Perak Branch, Tapah Campus, Malaysia. Her research interests are cybersecurity, big data analytics and machine learning. She is also Certified Computer Forensic Investigator. She can be contacted at email: samsi260@uitm.edu.my.

Noreen Izza Arshad is an Associate Professor with the Department of Computer and Information Sciences (CISD), Universiti Teknologi PETRONAS (UTP), Malaysia. Dr Noreen also heads the Positive Computing Research Cluster under the directive of the Institute of Autonomous Systems at the same university. Furthermore, she is lectures courses associated to Knowledge Management and Information Systems. She researches interest is tailored to understand the ways technology is being used, adapted, and adopted, and how it brings value to the society, business or organizations. She can be contacted at email: noreenizza@utp.edu.my.

Mohammad Nasir Abdullah is a senior lecturer in School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM). With a first degree in statistics from UiTM under his belt, Dr. Nasir ventured into the professional world as a data analysis and a business analyst for both multinarional and local companies. Fuelled by a passion for numbers, Dr. Nasir continued his academic pursuit, earning his second degree from USM and a certification as a data science specialist. Dr. Nasir was awarded his Ph.D. in Statistics from UiTM, cementing his place as a respected and knowledgeable authority in his field. His current research interest lies in ML classifiers in data classification. He can be contacted at email: nasir916@uitm.edu.my.